

501.43126X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): OGASAWARA, et al.
Serial No.: Not assigned
Filed: September 30, 2003
Title: STORAGE DEVICE SYSTEM AND STORAGE DEVICE SYSTEM
ACTIVATING METHOD
Group: Not assigned

LETTER CLAIMING RIGHT OF PRIORITY

Mail Stop Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

September 30, 2003

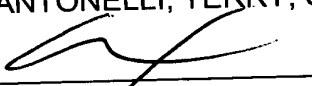
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Application No.(s) 2003-015525 filed January 24, 2003.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Carl I. Brundidge
Registration No. 29,621

CIB/amr
Attachment
(703) 312-6600

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日
Date of Application:

2003年 1月24日

出 願 番 号
Application Number:

特願2003-015525

[ST.10/C]:

[JP2003-015525]

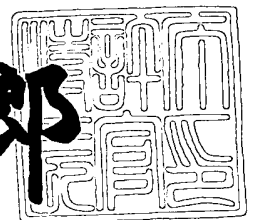
出 願 人
Applicant(s):

株式会社日立製作所

2003年 6月 9日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3044536

【書類名】 特許願

【整理番号】 K03001621A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 RAIDシステム事業部内

 【氏名】 小笠原 裕

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 RAIDシステム事業部内

 【氏名】 ▲高▼田 豊

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 RAIDシステム事業部内

 【氏名】 小林 直孝

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

 【識別番号】 100075096

 【弁理士】

 【氏名又は名称】 作田 康夫

【手数料の表示】

 【予納台帳番号】 013088

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

【物件名】 要約書 1
【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶装置システム、及び記憶装置システムの起動方法

【特許請求の範囲】

【請求項 1】

情報を格納する複数の記憶装置と、

前記複数の記憶装置に対する情報の格納を制御する記憶装置制御部と、

前記記憶装置制御部に接続される接続部と、

前記接続部を介して前記記憶装置制御部に接続されるとともに、自記憶装置システムの外部の第一のネットワークに接続され、前記外部の第一のネットワークを介して受けた第一の形式の情報を第二の形式の情報に変換して前記複数の記憶装置へのアクセスを要求される第一のプロセッサと、前記第一のプロセッサからのアクセス要求に応じて前記接続部及び前記記憶装置制御部を介して前記複数の記憶装置へアクセスするとともに、前記第一のプロセッサの起動を制御する第二のプロセッサとを有する第一の通信制御部とを有することを特徴とする記憶装置システム。

【請求項 2】

請求項 1 に記載の記憶装置システムにおいて、

前記記憶装置システムの外部の第二のネットワークに接続される第二の通信制御部と、

前記第一の通信制御部は、前記第二の通信制御部と同様の回路基板によって構成されていることを特徴とする記憶装置システム。

【請求項 3】

請求項 2 に記載の記憶装置システムにおいて、

前記第一のプロセッサは、自プロセッサのハードウェア診断を行うものであり

、
前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサのハードウェア診断の開始を要求するものであることを特徴とする記憶装置システム。

【請求項 4】

請求項 3 に記載の記憶装置システムにおいて、

前記第一の通信制御部及び前記第二の通信制御部に接続される管理端末とを有し、

前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサの第一の処理開始を要求するものであり、

前記第一のプロセッサは、前記第二のプロセッサからの前記第一の処理開始の要求に応じて、前記管理端末から第一のソフトウェアを取得することを特徴とする記憶装置システム。

【請求項 5】

請求項 4 に記載の記憶装置システムにおいて、

前記第一のプロセッサは、前記管理端末から取得した第一のソフトウェアによる制御に応じて、前記管理端末から第二のソフトウェアを取得して、前記複数の記憶装置に対して、前記接続部及び前記記憶装置制御部を介して書き込むことを特徴とする記憶装置システム。

【請求項 6】

請求項 5 に記載の記憶装置システムにおいて、

前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサの第二の処理開始を要求するものであり、

前記第一のプロセッサは、前記第二のプロセッサからの前記第二の処理開始の要求に応じて、前記複数の記憶装置に書き込まれた第二のソフトウェアを、前記接続部及び前記記憶装置制御部を介して取得することを特徴とする記憶装置システム。

【請求項 7】

請求項 4、6 に記載の記憶装置システムにおいて、

前記第一の処理開始要求及び前記第二の処理開始要求には、時刻情報が含まれることを特徴とする記憶装置システム。

【請求項 8】

請求項 1 に記載の記憶装置システムにおいて、

前記第一の通信制御部は、第三のソフトウェアを記憶する記憶装置を有し、

前記第一のプロセッサは、前記第一の通信制御部の起動に際して、前記第三のソフトウェアを起動させて、前記第二のプロセッサからの要求を待つものであることを特徴とする記憶装置システム。

【請求項 9】

請求項 8 に記載の記憶装置システムにおいて、
前記記憶装置システムの外部の第二のネットワークに接続される第二の通信制御部と、

前記第一の通信制御部は、前記第二の通信制御部と同様の回路基板によって構成されていることを特徴とする記憶装置システム。

【請求項 10】

請求項 9 に記載の記憶装置システムにおいて、

前記第一のプロセッサは、自プロセッサのハードウェア診断を行うものであり

、
前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサのハードウェア診断の開始を要求するものであることを特徴とする記憶装置システム。

【請求項 11】

請求項 10 に記載の記憶装置システムにおいて、

前記第一の通信制御部及び前記第二の通信制御部に接続される管理端末とを有し、

前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサの第一の処理開始を要求するものであり、

前記第一のプロセッサは、前記第二のプロセッサからの前記第一の処理開始の要求に応じて、前記管理端末から第一のソフトウェアを取得することを特徴とする記憶装置システム。

【請求項 12】

請求項 11 に記載の記憶装置システムにおいて、

前記第一のプロセッサは、前記管理端末から取得した第一のソフトウェアによる制御に応じて、前記管理端末から第二のソフトウェアを取得して、前記複数の

記憶装置に対して、前記接続部及び前記記憶装置制御部を介して書き込むことを特徴とする記憶装置システム。

【請求項 1 3】

請求項 1 2 に記載の記憶装置システムにおいて、

前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサの第二の処理開始を要求するものであり、

前記第一のプロセッサは、前記第二のプロセッサからの前記第二の処理開始の要求に応じて、前記複数の記憶装置に書き込まれた第二のソフトウェアを、前記接続部及び前記記憶装置制御部を介して取得することを特徴とする記憶装置システム。

【請求項 1 4】

請求項 1 1、1 3 に記載の記憶装置システムにおいて、

前記第一の処理開始要求及び前記第二の処理開始要求には、時刻情報が含まれることを特徴とする記憶装置システム。

【請求項 1 5】

情報を格納する複数の記憶装置と、前記複数の記憶装置に対する情報の格納を制御する記憶装置制御部と、前記記憶装置制御部に接続される接続部と、前記接続部を介して前記記憶装置制御部に接続されるとともに、自記憶装置システムの外部の第一のネットワークに接続される第一の通信制御部とを有する記憶装置システムの起動方法であって、

前記外部の第一のネットワークを介して受けた第一の形式の情報を第二の形式の情報に変換して前記複数の記憶装置へのアクセスを要求される第一のプロセッサは、前記第一のプロセッサからのアクセス要求に応じて前記接続部及び前記記憶装置制御部を介して前記複数の記憶装置へアクセスする第二のプロセッサによって、起動を制御されるものであり、

前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサのハードウェア診断の開始を要求し、

前記第一のプロセッサは、前記自プロセッサのハードウェア診断の開始要求に応じて、ハードウェア診断を行うものであることを特徴とする記憶装置システム

の起動方法。

【請求項 1 6】

請求項 1 5 に記載の記憶装置システムの起動方法において、

記憶装置システムは、前記第一の通信制御部及び前記第二の通信制御部に接続される管理端末とを有し、

前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサの第一の処理開始を要求し、

前記第一のプロセッサは、前記第二のプロセッサからの前記第一の処理開始の要求に応じて、前記管理端末から第一のソフトウェアを取得することを特徴とする記憶装置システムの起動方法。

【請求項 1 7】

請求項 1 6 に記載の記憶装置システムの起動方法において、

前記第一のプロセッサは、前記管理端末から取得した第一のソフトウェアによる制御に応じて、前記管理端末から第二のソフトウェアを取得して、前記複数の記憶装置に対して、前記接続部及び前記記憶装置制御部を介して書き込むことを特徴とする記憶装置システムの起動方法。

【請求項 1 8】

請求項 1 7 に記載の記憶装置システムの起動方法において、

前記第二のプロセッサは、前記第一のプロセッサに対して前記第一のプロセッサの第二の処理開始を要求し、

前記第一のプロセッサは、前記第二のプロセッサからの前記第二の処理開始の要求に応じて、前記複数の記憶装置に書き込まれた第二のソフトウェアを、前記接続部及び前記記憶装置制御部を介して取得することを特徴とする記憶装置システムの起動方法。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、複数の異種ネットワークに接続可能なように全く新たに発明された記憶装置システムに関し、特に記憶装置システムの起動を制御する方法に関する

【0002】

【従来の技術】

近年コンピュータシステムで取り扱われるデータ量が急激に増加している。かかる膨大なデータを効率よく利用し管理するために、複数のディスクアレイ装置（以下、記憶装置システムと称する）と情報処理装置とを専用のネットワーク（Storage Area Network、以下SANと記す）で接続し、記憶装置システムへの高速かつ大量なアクセスを実現する技術が開発されている。記憶装置システムと情報処理装置とをSANで接続し高速なデータ転送を実現するためには、ファイバチャネルプロトコルに従った通信機器を用いてネットワークを構築するのが一般的である。

【0003】

一方、複数の記憶装置システムと情報処理装置とをTCP/IP (Transmission Control Protocol/Internet Protocol) プロトコルを用いたネットワークで相互に接続し、記憶装置システムへのファイルレベルでのアクセスを実現する、NAS (Network Attached Storage) と呼ばれるネットワークシステムが開発されている。NASにおいては、記憶装置システムに対してファイルシステム機能を有する装置が接続されているため、情報処理装置からのファイルレベルでのアクセスが可能となっている。特に最近ではミッドレンジクラスやエンタープライズクラスと呼ばれるような、巨大な記憶資源を提供するRAID (Redundant Arrays of Inexpensive Disks) 方式で管理された記憶装置システムにファイルシステムを結合させた、大規模なNASが注目されている。

【0004】

【特許文献1】

特開2002-351703号公報

【0005】

【発明が解決しようとする課題】

しかしながら従来のNASは、TCP/IP通信機能及びファイルシステム機能を持たない記憶装置システムに、TCP/IP通信機能及びファイルシステム

機能を持った情報処理装置を接続させることにより実現されていた。そのため、上記接続される情報処理装置の設置スペースが必要であった。また上記情報処理装置と記憶装置システムとの間には、高速に通信を行う必要性から S A N で接続されていることが多く、そのための通信制御機器や通信制御機能を備える必要もあった。

【 0 0 0 6 】

本発明は上記課題を鑑みてなされたものであり、複数の異種ネットワークに接続可能なように全く新しく発明された記憶装置システム、及びかかる記憶装置システムを発明するにあたり必要とされる記憶デバイス制御装置、及びデバイス制御装置の起動を制御する方法を提供することを主たる目的とする。

【 0 0 0 7 】

【課題を解決するための手段】

本発明の記憶装置システムは、情報を格納する複数の記憶装置と、前記複数の記憶装置に対する情報の格納を制御する記憶装置制御部と、前記記憶装置制御部に接続される接続部とを有し、さらに、前記接続部を介して前記記憶装置制御部に接続されるとともに、自記憶装置システムの外部の第一のネットワークに接続され、前記外部の第一のネットワークを介して受けた第一の形式の情報を第二の形式の情報に変換して前記複数の記憶装置へのアクセスを要求される第一のプロセッサと、前記第一のプロセッサからのアクセス要求に応じて前記接続部及び前記記憶装置制御部を介して前記複数の記憶装置へアクセスするとともに、前記第一のプロセッサの起動を制御する第二のプロセッサとを有する第一の通信制御部とを有する。

【 0 0 0 8 】

【発明の実施の形態】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

【 0 0 0 9 】

まず、本実施の形態に係る記憶装置システムの全体構成を示すブロック図を図 1 に示す。

(全体構成例)

記憶装置システム600は、記憶デバイス制御装置100と記憶デバイス300とを備えている。記憶デバイス制御装置100は、情報処理装置200から受信したコマンドに従って記憶デバイス300に対する制御を行う。例えば情報処理装置200からデータの入出力要求を受信して、記憶デバイス300に記憶されているデータの入出力のための処理を行う。データは、記憶デバイス300が備えるディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリューム (Logical Unit) (以下、LUと記す) に記憶されている。また、記憶デバイス制御装置100は、情報処理装置200との間で、記憶装置システム600を管理するための各種コマンドの授受も行う。

【0010】

情報処理装置200はCPU (Central Processing Unit) やメモリを備えたコンピュータである。情報処理装置200が備えるCPUにより各種プログラムが実行されることによりさまざまな機能が実現される。情報処理装置200は、例えばパーソナルコンピュータやワークステーションであることもあるし、メインフレームコンピュータであることもある。

【0011】

図1において、情報処理装置1乃至3 (200) は、LAN (Local Area Network) 400を介して記憶デバイス制御装置100と接続されている。LAN400は、インターネットとすることもできるし、専用のネットワークとすることもできる。LAN400を介して行われる情報処理装置1乃至3 (200) と記憶デバイス制御装置100との間の通信は、例えばTCP/IPプロトコルに従って行われる。情報処理装置1乃至3 (200) からは、記憶装置システム600に対して、ファイル名指定によるデータアクセス要求 (ファイル単位でのデータ入出力要求。以下、ファイルアクセス要求と記す) が送信される。

【0012】

LAN400にはバックアップデバイス910が接続されている。バックアップデバイス910は具体的にはMOやCD-R、DVD-RAMなどのディスク系デバイス、DATテープ、カセットテープ、オープンテープ、カートリッジテープなどのテープ系デバイスである。バックアップデバイス910は、LAN400を介して記憶デバ

イス制御装置 1 0 0 との間で通信を行うことにより、記憶デバイス 3 0 0 に記憶されているデータのバックアップデータを記憶する。またバックアップデバイス 9 1 0 は情報処理装置 1 (2 0 0) と接続されるようにすることもできる。この場合は情報処理装置 1 (2 0 0) を介して記憶デバイス 3 0 0 に記憶されているデータのバックアップデータを取得するようにする。

【 0 0 1 3 】

記憶デバイス制御装置 1 0 0 は、チャンネル制御部 1 乃至 4 (1 1 0) を備える。記憶デバイス制御装置 1 0 0 は、チャンネル制御部 1 乃至 4 (1 1 0) により LAN 4 0 0 を介して情報処理装置 1 乃至 3 (2 0 0) からのファイルアクセス要求を個々に受け付ける。すなわち、チャンネル制御部 1 乃至 4 (1 1 0) には、個々に LAN 4 0 0 上のネットワークアドレス（例えば、IP アドレス）が割り当てられていてそれぞれが個別に NAS として振る舞い、個々の NAS があたかも独立した NAS が存在するかのように、NAS としてのサービスを情報処理装置 1 乃至 3 (2 0 0) に提供することができる。以下、チャンネル制御部 1 乃至 4 (1 1 0) を CHN と記す。このように 1 台の記憶装置システム 6 0 0 に個別に NAS としてのサービスを提供するチャンネル制御部 1 乃至 4 (1 1 0) を備えるように構成したことで、従来、独立したコンピュータで個々に運用されていた NAS サーバが一台の記憶システム 6 0 0 に集約される。そして、これにより記憶装置システム 6 0 0 の統括的な管理が可能となり、各種設定・制御や生涯管理、バージョン管理といった保守業務の効率化が図られる。

【 0 0 1 4 】

なお、本実施の形態に係る記憶デバイス制御装置 1 0 0 のチャンネル制御部 1 乃至 4 (1 1 0) は、後述するように、一体的にユニット化された回路基板上に形成されたハードウェアおよびこのハードウェアにより実行されるオペレーティングシステム（以下、OS と記す）やこの OS 上で動作するアプリケーションプログラム、あるいはこのハードウェアにより実行される実行可能オブジェクトコードなどのソフトウェアにより実現される。このように本実施例の記憶装置システム 6 0 0 では、従来ハードウェアの一部として実装されてきた機能がソフトウェアにより実現されている。このため、本実施例の記憶装置システム 6 0 0 では柔軟性

に富んだシステム運用が可能となり、多様で変化の激しいユーザニーズによりきめ細かなサービスを提供することが可能となる。

【0015】

情報処理装置3乃至4(200)はSAN(Storage Area Network)500を介して記憶デバイス制御装置100と接続されている。SAN500は、記憶デバイス300が提供する記憶領域におけるデータの管理単位であるブロックを単位として情報処理装置3乃至4(200)との間でデータの授受を行うためのネットワークである。SAN500を介して行われる情報処理装置3乃至4(200)と記憶デバイス制御装置100との間の通信は、一般にファイバチャネルプロトコルに従って行われる。情報処理装置3乃至4からは、記憶装置システム600に対して、ファイバチャネルプロトコルに従ってブロック単位のデータアクセス要求(以下、ブロックアクセス要求と記す)が送信される。

【0016】

SAN500にはSAN対応のバックアップデバイス900が接続されている。SAN対応バックアップデバイス900は、SAN500を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。

【0017】

記憶デバイス制御装置5(200)は、LAN400やSAN500等のネットワークを介さずに記憶デバイス制御装置100と接続されている。情報処理装置5(200)としては例えばメインフレームコンピュータとすることができる。情報処理装置5(200)と記憶デバイス制御装置100との間の通信は、例えばFICON(Fibre Connection)(登録商標)やESCON(Enterprise System Connection)(登録商標)、ACONARC(Advanced Connection Architecture)(登録商標)、FIBARC(Fibre Connection Architecture)(登録商標)などの通信プロトコルに従って行われる。情報処理装置5(200)からは、記憶装置システム600に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。

【0018】

記憶デバイス制御装置100は、チャネル制御部7乃至8(110)により情

報処理装置 5 (200) との間で通信を行う。以下、チャネル制御部 7 乃至 8 (110) をCHAと記す。

【0019】

SAN 500 には記憶装置システム 600 の設置場所 (プライマリサイト) とは遠隔した場所 (セカンダリサイト) に設置される他の記憶装置システム 610 が接続している。記憶装置システム 610 は、後述するレプリケーション又はリモートコピーの機能におけるデータの複製先の装置として利用される。なお、記憶装置システム 610 はSAN 500 以外にもATMなどの通信回線により記憶装置システム 600 に接続していることもある。この場合には例えばチャネル制御部 110 として上記通信回線を利用するためのインタフェース (チャネルエクステンダ) を備えるチャネル制御部 110 が採用される。

(記憶デバイス)

記憶デバイス 300 は、多数のディスクドライブ (物理ディスク) を備えており、情報処理装置 200 に対して記憶領域を提供する。データは、ディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域であるLUに記憶されている。ディスクドライブとしては、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等さまざまなものを用いることができる。なお、記憶デバイス 300 は例えば複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置 200 に対して提供される記憶領域は、RAIDにより管理された複数のディスクドライブにより提供されるようにすることもできる。

【0020】

記憶デバイス制御装置 100 と記憶デバイス 300 との間は図 1 のように直接に接続される形態とすることもできるし、ネットワークを介して接続するようにすることもできる。さらに記憶デバイス 300 は記憶デバイス制御装置 100 と一体として構成されることもできる。

【0021】

記憶デバイス 300 に設定されるLUには、情報処理装置 200 からアクセス可能なユーザLUや、チャネル制御部 110 の制御のために使用されるシステム

LU等がある。システムLUにはCHN110で実行されるOSも格納される。また各LUにはチャネル制御部110が対応付けられている。これによりチャネル制御部110ごとにアクセス可能なLUが割り当てられている。また上記対応付けは、複数のチャネル制御部110で一つのLUを共有するようにすることもできる。なお以下において、ユーザLUやシステムLUをそれぞれユーザディスク、システムディスク等とも記す。

(記憶デバイス制御装置)

記憶デバイス制御装置100は、チャネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140、管理端末160及び接続部150を備える。

【0022】

チャネル制御部110は、情報処理装置200との間で通信を行うための通信インタフェースを備え、情報処理装置200との間でデータ入出力コマンド等を授受する機能を備える。例えばCHN110は情報処理装置1乃至3(200)からのファイルアクセス要求を受け付ける。これによる記憶装置システム600はNASとしてのサービスを情報処理装置1乃至3(200)に提供することができる。またCHF110は情報処理装置3乃至4(200)からのファイバチャネルプロトコルに従ったブロックアクセス要求を受け付ける。これにより記憶装置システム600は高速アクセス可能なデータ記憶サービスを情報処理装置3乃至4(200)に対して提供することができる。またCHA110は情報処理装置5(200)からのFICONやESCON、ACONARC、FIBERC等のプロトコルに従ったブロックアクセス要求を受け付ける。これにより記憶装置システム600は情報処理装置5(200)のようなメインフレームコンピュータに対してもデータ記憶サービスを提供することができる。

【0023】

各チャネル制御部110は、管理端末160とともに内部LAN151等の通信網で接続されている。これによりチャネル制御部110に実行させるマイクロプログラム等を管理端末160から送信しインストールすることが可能となっている。チャネル制御部110の構成については後述する。

【0024】

接続部150はチャンネル制御部110、共有メモリ120、キャッシュメモリ130及びディスク制御部140と接続されている。チャンネル制御部110、共有メモリ120、キャッシュメモリ130及びディスク制御部140間でのデータやコマンドの授受は、接続部150を介することにより行われる。接続部150は、例えば高速スイッチングによりデータ伝送を行う超高速クロスバスイッチなどのスイッチ、又はバス等で構成される。チャンネル制御部110同士がスイッチで接続されていることで、個々のコンピュータ上で動作するNASサーバがLANを通じて接続する従来の構成に比べてチャンネル制御部110間の通信パフォーマンスが大幅に向上している。また、これにより高速なファイル共有機能や高速フェイルオーバーなどが可能となる。

【0025】

共有メモリ120およびキャッシュメモリ130は、チャンネル制御部110、ディスク制御部140により共有される記憶メモリである。共有メモリ120は主に制御情報やコマンド等を記憶する為に利用されるのに対し、キャッシュメモリ130は主にデータを記憶するために利用される。

【0026】

例えば、あるチャンネル制御部110が情報処理装置200から受信したデータ入出力コマンドが書き込みコマンドであった場合には、当該チャンネル制御部110は、書き込みコマンドを共有メモリ120に書き込むとともに、情報処理装置200から受信した書き込みデータをキャッシュメモリ130に書き込む。一方、ディスク制御部140は共有メモリ120を監視しており、共有メモリ120に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ130から書き込みデータを読み出して記憶デバイス300に書き込む。また、例えば、あるチャンネル制御部110が情報処理装置200から受信したデータ入出力コマンドが読み出しコマンドであった場合には、当該チャンネル制御部110は、読み出しコマンドを共有メモリ120に書き込むとともに、情報処理装置200から読み出しコマンドによって要求されたデータをキャッシュメモリ130から読み出す。仮に読み出しコマンドによって要求されたデー

タがキャッシュメモリ130に書き込まれていなかった場合、チャンネル制御部110又はディスク制御部140は、読み出しコマンドによって要求されたデータを記憶デバイス300から読み出して、キャッシュメモリ130に書き込む。

【0027】

なお、上記の本実施の形態においては、共有メモリ120及びキャッシュメモリ130がチャンネル制御部110及びディスク制御部140に対して独立に設けられていることについて記載されているが、本実施の形態はこの場合に限られるものでなく、共有メモリ120又はキャッシュメモリ130がチャンネル制御部110及びディスク制御部140の各々に分散されて設けられることも好ましい。この場合、接続部150は、分散された共有メモリ又はキャッシュメモリを有するチャンネル制御部110及びディスク制御部140を相互に接続させることになる。

【0028】

ディスク制御部140は、記憶デバイス300の制御を行う。例えば上述のように、チャンネル制御部110が情報処理装置200から受信したデータ書き込みコマンドに従って記憶デバイス300へデータの書き込みを行う。また、チャンネル制御部110により送信された論理アドレス指定によるLUへのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。記憶デバイス300における物理ディスクがRAIDにより管理されている場合には、RAID構成に従ったデータのアクセスを行う。またディスク制御部140は、記憶デバイス300に記憶されたデータの複製管理の制御やバックアップ制御を行う。さらにディスク制御部140は、災害発生時のデータ消失防止（ディザスタリカバリ）などを目的として、プライマリサイトの記憶装置システム600のデータの複製をセカンダリサイトに設置された他の記憶装置システム610にも記憶する制御（レプリケーション機能、またはリモートコピー機能）なども行う。

【0029】

各ディスク制御部140は管理端末160とともに内部LAN151等の通信網で接続されており、相互に通信を行うことが可能である。これにより、ディス

ク制御部140に実行させるマイクロプログラム等を管理端末160から送信しインストールすることが可能となっている。ディスク制御部140の構成については後述する。

(管理端末)

管理端末160は記憶装置システム600を保守・管理するためのコンピュータである。管理端末160を操作することにより、例えば記憶デバイス300内の物理ディスク構成の設定や、LUの設定、チャネル制御部110において実行されるマイクロプログラムのインストール等を行うことができる。ここで、記憶デバイス300内の物理ディスク構成の設定としては、例えば物理ディスクの増設や減設、RAID構成の変更(RAID1からRAID5への変更等)等を行うことができる。さらに管理端末160からは、記憶装置システム600の動作状態の確認や故障部位の特定、チャネル制御部110で実行されるOSのインストール等の作業を行うこともできる。また管理端末160はLANや電話回線等で外部保守センタと接続されており、管理端末160を利用して記憶装置システム600の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は例えばOSやアプリケーションプログラム、ドライバソフトウェアなどから通知される。この通知はHTTPプロトコルやSNMP(Simple Network Management Protocol)、電子メールなどにより行われる。これらの設定や制御は、管理端末160で動作するWebサーバが提供するWebページをユーザインタフェースとしてオペレータなどにより行われる。オペレータ等は、管理端末160を操作して障害監視する対象や内容の設定、障害通知先の設定などを行うこともできる。

【0030】

管理端末160は記憶デバイス制御装置100に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また管理端末160は、記憶デバイス制御装置100及び記憶デバイス300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

【0031】

管理端末160の構成を示すブロック図を図2に示す。

【0032】

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読み取り装置164、入力装置165、出力装置166及び記憶装置168を備える。

【0033】

CPU161は、管理端末160の全体の制御を司るもので、メモリ162に格納されたプログラム162cを実行することにより上記Webサーバとしての機能等を実現する。メモリ162には、物理ディスク管理テーブル162aとLU管理テーブル162bとプログラム162cとが記憶されている。

【0034】

物理ディスク管理テーブル162aは、記憶デバイス300に備えられる物理ディスク（ディスクドライブ）を管理するためのテーブルである。物理ディスク管理テーブル162aを図3に示す。図3においては、記憶デバイス300が備える多数の物理ディスクのうち、ディスク番号#001乃至#006までが示されている。それぞれの物理ディスクに対して、容量、RAID構成、使用状況が示されている。

【0035】

LU管理テーブル162bは、上記物理ディスク上に論理的に設定されるLUを管理するためのテーブルである。LU管理テーブル162bを図4に示す。図4においては、記憶デバイス300上に設定される多数のLUのうち、LU番号#1乃至#3までが示されている。それぞれのLUに対して、物理ディスク番号、容量、RAID構成が示されている。

【0036】

記憶媒体読取装置164は、記録媒体167に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ162や記憶装置168に格納される。従って、例えば記録媒体167に記録されたプログラム162cを、記録媒体読取装置164を用いて記録媒体167から読み取って、メモリ162や記憶装置168に格納するようにすることができる。

。記録媒体167としてはフレキシブルディスクやCD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置162は管理端末160に内蔵されている形態とすることもできる。記憶装置168は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等である。入力装置165は、オペレータ等による管理端末160へのデータ入力等のために用いられる。入力装置165としては例えばキーボードやマウス等が用いられる。出力装置166は、情報を外部に出力するための装置である。出力装置166としては例えばディスプレイやプリンタ等が用いられる。ポート163は内部LAN151に接続されており、これにより管理端末160はチャネル制御部110やディスク制御部140等と通信を行うことができる。またポート163は、LAN400に接続するようにすることもできるし、電話回線に接続するようにすることもできる。

(外観図)

次に、本実施の形態に係る記憶装置システム600の外観構成を図5に示す。
また、記憶デバイス制御装置100の外観構成を図6に示す。

【0037】

図5に示すように、本実施の形態に係る記憶装置システム600は記憶デバイス制御装置100および記憶デバイス300がそれぞれの筐体に収められた形態をしている。記憶デバイス制御装置100の筐体の両側に記憶デバイス300の筐体が配置されている。

【0038】

記憶デバイス制御装置100は、正面中央部に管理端末160が備えられている。管理端末160はカバーで覆われており、図6に示すようにカバーを開けることにより管理端末160を使用することができる。なお図6に示した管理端末160はいわゆるノート型パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

【0039】

管理端末160の下部には、チャネル制御部110を装着するためのスロットが設けられている。各スロットにはチャネル制御部110のボードが装着される。本実施の形態に係る記憶装置システム600においては、例えばスロットは8

つあり、図5および図6には8つのスロットにチャンネル制御部110を装着するためのガイドレールが設けられている。ガイドレールに沿ってチャンネル制御部110をスロットに挿入することにより、チャンネル制御部110を記憶デバイス制御装置100に装着することができる。また各スロットに装着されたチャンネル制御部110は、ガイドレールに沿って手前方向に引き抜くことにより取り外すことができる。また各スロットの奥手方向正面部には、各チャンネル制御部110を記憶デバイス制御装置100と電氣的に接続するためのコネクタが設けられている。チャンネル制御部110には、CHN、CHF、CHAがあるが、いずれのチャンネル制御部110もサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせているため、8つのスロットにはいずれのチャンネル制御部110も装着することが可能である。従って、例えば8つのスロット全てにCHN110を装着するようにすることもできる。また例えば図1に示したように、4枚のCHN110と、2枚のCHF110と、2枚のCHA110とを装着するようにすることもできる。チャンネル制御部110を装着しないスロットを設けることもできる。

【0040】

なお、上述したように、チャンネル制御部110は上記各スロットに装着可能なボード、すなわち同一のユニットに形成された一つのユニットとして提供されるが、上記同一のユニットは複数枚数の基板から構成されているようにすることもできる。つまり、複数枚数の基板から構成されていても、各基板が相互に接続されて一つのユニットとして構成され、記憶デバイス制御装置100のスロットに対して一体的に装着できる場合は、同一の回路基板の概念に含まれる。

【0041】

ディスク制御部140や共有メモリ120等の、記憶デバイス制御装置100を構成する他の装置については図5および図6には示されていないが、記憶デバイス制御装置100の背面側当に装着されている。

【0042】

また記憶デバイス制御装置100には、チャンネル制御部110とらおから発生する熱を放出するためのファン170が設けられている。ファン170は記憶デ

バイス制御装置 1 0 0 の上面部に設けられるほか、チャンネル制御部 1 1 0 用スロットの上部にも設けられている。

【 0 0 4 3 】

ところで、筐体に収容されて構成される記憶デバイス制御装置 1 0 0 および記憶デバイス 3 0 0 としては、例えば S A N 製品として製品化されている従来構成の装置を利用することができる。特に上記のように C H N のコネクタ形状を従来構成の筐体に設けられているスロットにそのまま装着できる形状とすることとで従来構成の装置をより簡単に利用することができる。つまり、本実施例の記憶装置システム 6 0 0 は、既存の製品を利用することで容易に構築することができる。

【 0 0 4 4 】

さらに、本実施の形態によれば、記憶装置システム 6 0 0 内に C H N 1 1 0、C H F 1 1 0、C H A 1 1 0 を混在させて装着させることにより、異種ネットワークに接続される記憶装置システムを実現できる。具体的には、記憶装置システム 6 0 0 は、C H N 1 1 0 を用いて L A N 1 4 0 に接続し、かつ C H F 1 1 0 を用いて S A N 5 0 0 に接続するという、S A N - N A S 統合記憶装置システムである。

(チャンネル制御部)

本実施の形態に係る記憶装置システム 6 0 0 は、上述の通り C H N 1 1 0 により情報処理装置 1 乃至 3 (2 0 0) からのファイルアクセス要求を受け付け、N A S としてのサービスを情報処理装置 1 乃至 3 (2 0 0) に提供する。

【 0 0 4 5 】

C H N 1 1 0 のハードウェア構成を図 7 に示す。この図に示すように C H N 1 1 0 のハードウェアは一つのユニットで構成される。以下、このユニットのことを N A S ボードと記す。N A S ボードは一枚もしくは複数枚の回路基板を含んで構成される。より具体的には、N A S ボードは、ネットワークインタフェース部 1 1 1、入出力制御部 1 1 4、ボード接続用コネクタ 1 1 6、通信コネクタ 1 1 7 及びファイルサーバ部 8 0 0 を備え、これらが同一のユニットに形成されて構成されている。さらに、入出力制御部 1 1 4 は、N V R A M (Non Volatile RAM

） 1 1 5 及び I / O (Input/Output) プロセッサ 1 1 9 を有する。

【 0 0 4 6 】

ネットワークインタフェース部 1 1 1 は、情報処理装置 2 0 0 との間で通信を行うための通信インタフェースを備えている。CHN 1 1 0 の場合は、例えば TCP / IP プロトコルに従って情報処理装置 2 0 0 から送信されたファイルアクセス要求を受信する。通信コネクタ 1 1 7 は、情報処理装置 2 0 0 との間で通信を行うためのコネクタである。CHN 1 1 0 の場合は、LAN 4 0 0 に接続可能なコネクタであり、例えばイーサネット（登録商標）に対応している。

【 0 0 4 7 】

ファイルサーバ部 8 0 0 は、CPU 1 1 2、メモリ 1 1 3、BIOS (Basic Input/Output System) 8 0 1 及び NVRAM 8 0 4 を有する。CPU 1 1 2 は、CHN 1 1 0 を NAS ボードとして機能させるための制御を司る。CPU 1 1 2 は、NFS 又は CIFS 等のファイル共有プロトコル及び TCP / IP の制御、ファイル指定されたファイルアクセス要求の解析、メモリ 1 1 3 内の制御情報へのファイル単位のデータと記憶デバイス 3 0 0 内の LU との変換テーブル（図示せず）を用いた相互変換、記憶デバイス 3 0 0 内の LU に対するデータ書き込み又は読み出し要求の生成、データ書き込み又は読み出し要求の I / O プロセッサ 1 1 9 への送信等処理する。BIOS 8 0 1 は、例えば CHN 1 1 0 に電源が投入された際に、CPU 1 1 2 を起動する過程で最初にメモリ 1 1 3 にロードされ実行されるソフトウェアであり、例えばフラッシュメモリなどの不揮発性の媒体に保存されて CHN 1 1 0 上に実装されている。CPU 1 1 2 は、BIOS 8 0 1 からメモリ 1 1 3 上に読み込まれたソフトウェアを実行することにより、CHN 2 1 上の CPU 1 1 2 が関係する部分の初期化、診断などを行うことができる。さらに、CPU 1 1 2 は、BIOS 8 0 1 から I / O プロセッサ 1 1 9 にコマンドなどの指示を発行することにより、記憶デバイス 3 0 0 から所定のプログラム、例えば OS のブート部などをメモリ 1 1 3 に読み込むことができる。読み込まれた OS のブート部は、さらに記憶デバイス 3 0 0 に格納されている OS の主要部分をメモリ 1 1 3 に読み込む動作をし、これにより CPU 1 1 2 上で OS が起動され、例えばファイルサーバとしての処理が実行できるようになる。ま

た、ファイルサーバ部 8 0 0 は、P X E (Preboot eXecution Environment) などの規約にしたがうネットワークブートローダを格納する N V R A M 8 0 4 を実装し、後述するネットワークブートを行わせることも可能である。

【 0 0 4 8 】

メモリ 1 1 3 にはさまざまなプログラムやデータが記憶される。例えば図 8 に示すメタデータ 7 3 0 やロックテーブル 7 2 0、また図 1 6 に示される NAS マネージャ 7 0 6 等の各種プログラムが記憶される。メタデータ 7 3 0 は、ファイルシステムが管理しているファイルに対応させて生成される情報である。メタデータ 7 3 0 には例えばファイルのデータが記憶されている LU 上のアドレスやデータサイズなど、ファイルの保管場所を特定するための情報が含まれる。メタデータ 7 3 0 にはファイルの容量、所有者、更新時刻等の情報が含まれることもある。また、メタデータ 7 3 0 はファイルだけでなくディレクトリに対応させて生成されることもある。メタデータ 7 3 0 の例を図 9 に示す。メタデータ 7 3 0 は記憶デバイス 3 0 0 上の各 LU にも記憶されている。

【 0 0 4 9 】

ロックテーブル 7 2 0 は、情報処理装置 1 乃至 3 (2 0 0) からのファイルアクセスに対して排他制御を行うためのテーブルである。排他制御を行うことにより情報処理装置 1 乃至 3 (2 0 0) でファイルを共用することができる。ロックテーブル 7 2 0 を図 1 0 に示す。図 1 0 に示すようにロックテーブル 7 2 0 には、ファイルロックテーブル 7 2 1 と LU ロックテーブル 7 2 2 とがある。ファイルロックテーブル 7 2 1 は、ファイルごとにロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置 2 0 0 によりあるファイルがオープンされている場合に当該ファイルにロックが掛けられる。ロックが掛けられたファイルに対する他の情報処理装置 2 0 0 によるアクセスは禁止される。LU ロックテーブル 7 2 2 は、LU ごとにロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置 2 0 0 により、ある LU に対するアクセスが行われている場合に当該 LU にロックが掛けられる。ロックが掛けられた LU に対する他の情報処理装置 2 0 0 によるアクセスは禁止される。

【 0 0 5 0 】

入出力制御部114は、ディスク制御部140キャッシュメモリ130、共有メモリ120及び管理端末160との間でデータやコマンドの授受を行う。入出力制御部114はI/Oプロセッサ119及びNVRAM115を備えている。I/Oプロセッサ119は例えば1チップのマイコンで構成される。I/Oプロセッサ119は、記憶デバイス300内のLUに対するデータ書き込み又は読み出し要求やデータの授受を制御し、CPU112とディスク制御部140との間の通信を中継する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0051】

図11は、CHN110上のCPU112とI/Oプロセッサ119との通信経路について具体的に示す。I/Oプロセッサ119と、CPU112は、CHN110上に実装された通信メモリ802、ハードウェアレジスタ群803で物理的に接続されている。通信メモリ802およびハードウェアレジスタ群803は、それぞれCPU112およびI/Oプロセッサ119のいずれからもアクセスが可能である。ハードウェアレジスタ群803は、CPU112に対して電源を投入又は切断する回路に接続される。これにより、I/Oプロセッサ119は、ハードウェアレジスタ群803にアクセスすることによって、ハードウェアレジスタ群803を介してCPU112の電源を操作することが可能となる。ハードウェアレジスタ群803は、必要に応じて、CPU112あるいはI/Oプロセッサ119がハードウェアレジスタ群803にアクセスを行った際に、アクセス対象の相手先に割り込み信号などを生成して、アクセスが行われたことを通知する等の複数の機能を有する。これら複数の機能は、ハードウェアレジスタ群803を構成する各レジスタにそれぞれハードウェア的に割り当てられる。

【0052】

CHN110上の通信メモリ802に格納されるデータ構造の例を、図12及び図13に示す。図12は、I/Oプロセッサ119からCPU112へ情報を受け渡すために使用されるデータ構造であり、図13は、CPU112からI/O

Ｏプロセッサ１１９へ情報を受け渡すために使用されるデータ構造である。ＣＰＵ１１２とＩ／Ｏプロセッサ１１９との間でやり取りされる情報は、主に、電源投入などを契機として、ＣＰＵ１１２及びＩ／Ｏプロセッサ１１９が起動する際に授受される情報群である。

【００５３】

Ｉ／Ｏプロセッサ１１９からＣＰＵ１１２に渡される情報としては、起動デバイス種別、診断実行フラグ、複数のドライブ番号、時刻情報、コマンドリトライ回数、コマンドタイムアウト値及び複数の温度情報がある。起動デバイス種別は、ＣＰＵ１１２が起動するときに、ＢＩＯＳ８０１の制御によって起動されるデバイスの種類であり、例えばネットワーク、ディスクドライブなどの種別がある。ドライブ番号は、起動デバイス種別がディスクドライブであったときに、ＯＳのロード元のディスクドライブを選択するための番号である。なお、本実施の形態においては、記憶デバイス３００内にＬＵという概念を有しており、ＬＵに対してＯＳ等が格納されているため、ＬＵ毎に区別されるＬＵ番号をドライブ番号と考える。ドライブ番号には例えば優先度が設けられており、仮にドライブ番号０がドライブ番号１に優先する場合、ＣＰＵ１１２は、まずドライブ番号０に指定されているＬＵからの起動を試みて、その起動が失敗した場合に、ドライブ番号１に指定されているＬＵからの起動を試みるというような動作が可能である。診断実行フラグは、ＣＰＵ１１２が起動する際に、Ｉ／Ｏプロセッサ１１９からＢＩＯＳ８０１に対して、ファイルサーバ部８００周辺のハードウェア診断を実行するか否かを指示することに利用される。例えば、ファイルサーバ部８００の初期化が完了した時点で、ＣＰＵ１１２のみを再起動したような場合には、ＢＩＯＳ８０１が再度ハードウェア診断を実行することを要しない。このような場合に、Ｉ／Ｏプロセッサ１１９が診断実行フラグを適宜設定することによって、ＣＰＵ１１２によるファイルサーバ部８００のハードウェア診断の再度実行を抑止できる。時刻情報は、ＣＰＵ１１２上でＢＩＯＳ１１２やＯＳが動作する際に使用される。Ｉ／Ｏプロセッサ１１４は、管理端末１６０から時刻情報を取得して、ＣＰＵ１１２に渡される。これにより、管理端末１６０、Ｉ／Ｏプロセッサ１１４及びＣＰＵ１１２は、これらの三者間で時刻情報の同期をとることが可能と

なる。コマンドリトライ回数、コマンドタイムアウト値は、CPU112からI/Oプロセッサ119に発行されたコマンドが失敗したときのCPU112上のBIOS801あるいはOSの動作、及びタイムアウトの条件等である。温度情報は、CPU112が自らの温度変化の異常を検知できるようにするために、CPU112に対して設定される値である。

【0054】

このように、本実施の形態によれば、I/Oプロセッサ114が、起動デバイス種別、ドライブ番号、時刻情報、コマンドリトライ回数、コマンドタイムアウト値及び複数の温度情報などの値を自由に設定することができる。なお、本実施の形態はこの場合に限られることなく、これらの値がBIOSの不揮発性メモリ内に初期値として格納されることも好ましい。また、これらの値がオペレータによって管理端末160から入力されることにより、又はこれらの値が管理端末160のメモリ上に予め登録されることにより、管理端末160からI/Oプロセッサ114に対して渡されることとすることも好ましい。診断実行フラグは、I/Oプロセッサ114の起動中の論理的な判断によって設定され、又はオペレータによって設定される。I/Oプロセッサ114の起動中の論理的な判断によって診断実行フラグが設定される場合には、CPU112の動作、又はCPU112上にロードされて動作するBIOS801の動作をI/Oプロセッサ119から制御することが可能である。

【0055】

図13は、CPU112からI/Oプロセッサ119へ情報を受け渡すためのデータ構造である。BIOSバージョンは、BIOS801のオブジェクトコードのバージョンであり、CPU112からI/Oプロセッサ119に渡され、さらにI/Oプロセッサ119から管理端末160に渡される。MACアドレスは、CPU112のMACアドレスである。MACアドレスは、ハードウェア的に世界で一意的な識別子であり、IPプロトコルによってIPアドレスをLAN上のDHCPサーバに割り当てる時などに必要な情報である。なお、0パディング情報はワード境界を埋めるためのものであり、情報とは無関係である。

【0056】

図14は、CPU112とI/Oプロセッサ119とを、内部LAN151によって接続しているハードウェア構成図である。このように、CPU112とI/Oプロセッサ119とは、ともに内部LAN151によっても接続されており、内部LAN151を介して管理端末160との通信が可能である。これにより、例えば、CPU112は、NVRAM804に予め格納されているネットワークブートローダを実行することにより、管理端末160から起動用のソフトウェアをメモリ113にダウンロードし、起動用のソフトウェアを実行することができる。これによって例えば、管理端末160をサーバとし、CPU112をクライアントとするネットワークブートプロセスが実行される。なお、ネットワークブートは、例えばPXEなどの規約に従い、クライアント上のネットワークブートローダと管理端末160上で動作するサーバとが、IPプロトコル、DHCP、TFTP、FTPなどのプロトコルを組み合わせることにより、LAN上の管理端末160に存在するOSのブートイメージを起動及び実行する方法である。

【0057】

図15は、ディスク制御部140のハードウェア構成を示すブロック図である。既に述べた通り、ディスク制御部は、記憶デバイス300に接続されるとともに接続部150を介してCHN112に接続され、ディスク制御部140独自で、又はCHN112によって制御されることにより、記憶デバイス300に対してデータの読み書きを行う。

【0058】

ディスク制御部140は、インタフェース部141、メモリ143、CPU142、NVRAM144及びボード接続用コネクタ145を備え、これらが一体的なユニットとして形成されている。

【0059】

インタフェース部141は、接続部150を介してチャネル制御部110等と通信を行うための通信インタフェース、記憶デバイス300と通信を行うための通信インタフェース、内部LAN151を介して管理端末160と通信を行うための通信インタフェースを備えている。

【0060】

CPU142は、ディスク制御部140全体の制御を司るとともに、チャネル制御部110や記憶デバイス300、管理端末160との間の通信を行う。メモリ143やNVRAM144に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部140の機能が実現される。ディスク制御部140により実現される機能としては、記憶デバイス300の制御やRAID制御、記憶デバイス300に記憶されたデータの複製管理やバックアップ制御、リモートコピー制御等である。

【0061】

NVRAM144はCPU142の制御を司るプログラムを格納する不揮発性メモリである。NVRAM144に記憶されるプログラムの内容は、管理端末160や、NASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0062】

またディスク制御部140はボード接続用コネクタ145を備えている。ボード接続用コネクタ145が記憶デバイス制御装置100側のコネクタと接続することにより、ディスク制御部140は、記憶デバイス制御装置100と電氣的に接続される。

(ソフトウェア構成図)

図16は、本実施の形態に係る記憶装置システム600におけるソフトウェア構成図である。既に述べたように、CHN110上には、CPU112およびI/Oプロセッサ119が存在する。CPU112およびI/Oプロセッサ119は、それぞれ1つつづつであってもよいし、それぞれ複数存在してもよい。CPU112上では、OS701とNASマネージャ706等の多様なアプリケーションとが実行されることにより、CPU112はNASサーバとして動作する。I/Oプロセッサ119上では、コントローラとしてのマイクロプログラムが動作している。ディスク制御部140では、RAID制御部740がCPU142上で動作している。管理端末160の上では、CPU161がネットブートサーバ703として動作する。ネットブートサーバ703は、記録媒体167又は記憶装置168等から内部LAN151を介して、ミニカーネル704、OSイメージ705等をCHN110上

のCPU112に転送する。ネットブートサーバ703は、例えば、DHCP (Dynamic Host Configuration Protocol) サーバなどを有し、CPU112、CPU161及びI/Oプロセッサ119にIPアドレス又はMACアドレスを割り当てる等して、管理端末160とCPU112、CPU161及びI/Oプロセッサ119との間の転送を行う。ネットブートを行うとき、例えば、CPU112は、クライアントとして、ネットブートサーバ703に対してDHCP要求及びファイル転送要求等を要求する。CPU112は、ネットブートの手順を経て、CPU112上でミニカーネル704を動作させることになる。最終的に、CPU112は、I/Oプロセッサ119を経由してOSイメージ705を記憶デバイス300にインストールさせる。

【0063】

なお、図16は、情報処理装置200のソフトウェア構成についても明示してある。情報処理装置200は、NFS (Network File System) 711を有するもの、又はCIFS (Common Internet File System) 713を有するものが存在する。NFS 711は、主にUNIX (登録商標) 系のオペレーティングシステム714によって用いられるファイル共有プロトコルであり、CIFS 713は、主にWindows (登録商標) 系のOS 715によって用いられるファイル共有プロトコルである。

(記憶装置システムの起動処理及びインストール処理)

CHN110に電源を投入した際に、CHN110等が起動する手順について説明する。起動手順を経ることにより、CHN110は、NASサーバとして動作することとなる。CHN110に電源を投入するためには、例えば予め記憶装置システム600に対してCHN110を挿入した後に、記憶装置システム600自体の電源を投入する方法がある。また、既に電源の投入されている記憶装置システム600に対し、あらたにCHN110を挿入することにより、CHN110の電源回路に記憶装置システム600から給電を行い、CHN110に電源を投入する方法もある。ここでは前者の、予め記憶装置システム600に対してCHN110を挿入した後に、記憶装置システム600自体の電源を投入する方法について説明する。

【 0 0 6 4 】

図 1 7 は、記憶装置システム 6 0 0 に電源を投入し、OS イメージ 7 0 5 を管理端末 1 6 0 から記憶デバイス 3 0 0 にインストールし、記憶デバイス 3 0 0 にインストールされた OS から起動して NAS システムとして動作するまでの流れが示されたものである。図 1 7 は、CPU 1 1 2、I/O プロセッサ 1 1 9、管理端末 1 6 0 及び記憶デバイス 3 0 0 を関連づけながら示されたものである。図の上から下に向かって時間が経過してゆく。図中において、実線の矢印は制御情報の伝達を示し、点線の矢印は OS ファイルなどのデータの流れを示す。

【 0 0 6 5 】

以下、順を追って説明する。

【 0 0 6 6 】

はじめに記憶装置システム 6 0 0 に電源が投入される（CPU 1 1 2 ステップ 1、I/O プロセッサ 1 1 9 ステップ 1、管理端末 1 6 0 ステップ 1、記憶デバイス 3 0 0 ステップ 1）。電源の投入に応じて、CPU 1 1 2 は、自動的に BIOS の起動を始める（CPU 1 1 2 ステップ 2）。I/O プロセッサ 1 1 9 は、CHN 1 1 0 内のハードウェアの初期化を行う（I/O プロセッサ 1 1 9 ステップ 2）。管理端末 1 6 0 では、端末上の OS およびソフトウェアが起動される（管理端末 1 6 0 ステップ 2）。記憶デバイス 3 0 0 では、ディスクドライブの起動が行われ、スピニング動作が行われる（記憶デバイス 3 0 0 ステップ 2）。ディスクドライブのスピニングの際には物理的な回転運動とその安定、ハードウェア診断などが行われるため、ある程度の時間を要する。ここでは、およそ CPU 1 1 2 が二度目の起動を行った時期にスピニングが完了するように制御している（記憶デバイス 3 0 0 ステップ 8）。

【 0 0 6 7 】

CPU 1 1 2 上では、BIOS 8 0 1 が I/O プロセッサ 1 1 9 からの指示を待つ状態になっている（CPU 1 1 2 ステップ 3）。I/O プロセッサ 1 1 9 は、自身の初期化及び CHN 1 1 0 内のハードウェアの初期化が終了すると、CPU 1 1 2 に対して、診断開始要求を発行することにより、ハードウェア診断の開始を指示する（I/O プロセッサ 1 1 9 ステップ 4）。このとき、診断開始

要求は、通信メモリ802あるいはハードウェアレジスタ群803、又はその双方を用いて伝達される。

【0068】

CPU112は、診断開始要求を検出すると、ハードウェアの診断を開始する(CPU112-ステップ4)。ハードウェアの診断が終了した場合(CPU112-ステップ5)、CPU112は、自身に設けられた内部LAN151用ポートに予め設定されているMACアドレスを通信メモリ802に格納する(CPU112-ステップ6)。I/Oプロセッサ119は、MACアドレスが通信メモリ802に格納されたことを検知すると(I/Oプロセッサ119-ステップ6)、ハードウェアレジスタ群803にアクセスすることによりCPUリセット指示を発行し、を行い、CPU112を物理的にリセットさせる(I/Oプロセッサ119-ステップ7)。これにより、CPU112は、一次的に電源が落とされた後、電源が再投入された状態となる(CPU112-ステップ7)。CPU112は、再度BIOS起動動作を行い(CPU112-ステップ8)、再びI/Oプロセッサ119からの指示を待つ状態になる(CPU112-ステップ9)。

【0069】

管理端末160は、起動を完了した場合、インストール操作を開始する。インストール操作は、オペレータによって行われることも良いし、予め設定されたプログラムに従ってCPU161の制御によって実行されることも良い。以下においては、インストール操作はオペレータによって行われ、管理端末160上のCPU161がオペレータの入力装置165に対する入力に応じた処理を行うこととする。インストール操作が開始された後、管理端末160は、MACアドレスをI/Oプロセッサ119に対して問い合わせ、I/Oプロセッサ119からCPU112のMACアドレスを取得する(管理端末160-ステップ10)。オペレータは、OSが起動してからのコマンドリトライ回数、タイムアウト時間、温度情報などを入力し設定する。(管理端末160-ステップ11)。オペレータによって入力されたこれらの値は、管理端末160から内部LAN151を介してCHN110のI/Oプロセッサ119に送信される。また、オペレータ

は、管理端末160に対して、OSを新規にインストールするように指示を出す。オペレータが起動デバイス種別をネットブートと設定することにより、OSインストール指示は、管理端末160からI/Oプロセッサ119に伝達される（管理端末160-ステップ11）。

【0070】

I/Oプロセッサ119はOSインストール指示を受領すると、記憶デバイス300が使用可能になっているかどうかを判断するために、ドライブスピンアップ判定を行う（I/Oプロセッサ119-ステップ12）。記憶デバイス300は既にスピンアップ状態となっているので（記憶デバイス300-ステップ8）、I/Oプロセッサ119は、ドライブスピンアップ完了を検出する（I/Oプロセッサ119-ステップ13）。ここでドライブスピンアップ完了を検出できなかった場合、I/Oプロセッサ119は、管理端末160に対してエラー応答を行うことによって、ドライブがまだスピンアップしていないことを通知する。ドライブスピンアップ完了を検出した場合、I/Oプロセッサ119は、管理端末160から受領した温度情報、コマンドリトライ回数及びリトライタイムアウト時間等の値に加えて、診断実行フラグ及び起動デバイス種別を、通信メモリ802に格納する（I/Oプロセッサ119-ステップ14。ハードウェア診断については、すでに（CPU112-ステップ4）および（CPU112-ステップ5）で完了しているため、再度行う必要はない。したがって、I/Oプロセッサ119は、診断実行フラグに対してハードウェア診断のスキップを指示させる値を格納する。また、CPU112は、I/Oプロセッサ指示待ち（CPU112-ステップ9）後の起動において、管理端末160上に存在するOSイメージ705を記憶デバイス300にインストールする必要がある。したがって、I/Oプロセッサ119は、起動デバイス種別としてネットブートを指定して、通信メモリ802に格納する（I/Oプロセッサ119-ステップ14）。I/Oプロセッサ119は、通信メモリ802に種々の情報が格納されたら、ハードウェアレジスタ群803にアクセスして、情報が格納されたことをCPU112に対して通知する（I/Oプロセッサ119-ステップ15）。この通知は、CPU112に対して処理実行を指示する目的を持つ。

【0071】

CPU112は、I/Oプロセッサ119からの処理続行の指示を検知する（CPU112-ステップ15）。CPU112は、温度情報、コマンドリトライ回数、リトライタイムアウト時間、診断実行フラグ及び起動デバイス種別を通信メモリ802から取得する（CPU112-ステップ16）。起動デバイス種別がネットブートを指示しているため、CPU112は、ネットブート処理を開始する（CPU112-ステップ17）。ネットブートは、CPU112上で動作するブートローダと管理端末160上で動作するネットブートサーバ703とが相互に内部LAN151を通じてデータを通信することにより、進行される。ネットブートサーバ703は、（管理端末160-ステップ10）で取得されたCPU112のMACアドレス情報を用いて、通信する。CPU112は、PXEなどの規約に従って、管理端末160上のネットブートサーバ703に対してネットブート要求を発行する（CPU112-ステップ18）。管理端末160は、内部LAN151を介してネットブート要求を受領する（管理端末160-ステップ18）。ネットブート動作中において、CPU112はネットブートクライアントとして動作する（CPU112-ステップ19）。一方、管理端末160は、ネットブートサーバとして動作する（管理端末160-ステップ19）。ネットブート処理における通信の結果、管理端末160は、ミニカーネル704をCPU112上にダウンロードさせる。CPU112上にダウンロードされたミニカーネル704は、管理端末160上のOSイメージ705を記憶デバイス300に転送させる（CPU112-ステップ20）。この場合、OSイメージ705は、CHN110上のCPU112の制御に従って、I/Oプロセッサ119、キャッシュメモリ130及びディスク制御部140を介して、記憶デバイス300に対してインストールされる。インストール処理が完了した場合、CPU112は、インストール完了通知をI/Oプロセッサ119に対して発行する（CPU112-ステップ21）。I/Oプロセッサ119は、インストール完了通知を検出した場合（I/Oプロセッサ119-ステップ21）、ハードウェアレジスタ群803にアクセスすることによりCPUリセット指示を発行し、CPU112をリセットさせる（I/Oプロセッサ119-ステップ22）。これ

により、CPU112は、電源が再投入された状態になる（CPU112ーステップ22）。CPU112は、再度BIOS801を起動し（CPU112ーステップ23）、再びI/Oプロセッサ119からの指示を待つ状態になる（CPU112ーステップ24）。

【0072】

I/Oプロセッサ119は、時刻情報、起動デバイス種別及び起動ドライブ番号情報を、通信メモリ802に格納する（I/Oプロセッサ119ーステップ25）。この場合、I/Oプロセッサ119は、起動デバイス種別をディスクドライブと設定する。I/Oプロセッサ119は、ハードウェアレジスタ群803にアクセスすることにより、CPU112に対して処理続行指示を発行する（I/Oプロセッサ119ーステップ26）。

【0073】

CPU112は、CPU処理続行指示を検出した場合（CPU112ーステップ26）、通信メモリ802から時刻情報を取得して（CPU112ーステップ27）、I/Oプロセッサ119に対してコマンドを発行することによりディスクブートを開始させる（CPU112ーステップ28）。CPU112は、I/Oプロセッサ119を介して、記憶デバイス300に格納されているOSイメージ705をロードさせる（CPU112ーステップ29）。この場合、OSイメージ705は、ディスク制御部140、キャッシュメモリ130及びI/Oプロセッサ119を介して、記憶デバイス300からCPU112に対してロードされる。

【0074】

OSイメージ705がCPU112にロードされた場合、CPU112の処理がBIOS801からOSイメージ705に移り、OSが起動される（CPU112ーステップ30）。これにより、CHN110はNASとして動作できるようになり、記憶装置システム600はNASシステムとして起動できる（CPU112ーステップ31）。

【0075】

このように、本実施の形態によれば、記憶装置システム600の電源投入に際

して I/O プロセッサ 1 1 9 が CPU 1 1 2 を制御することにより、ネットワークを介した OS の新規インストール、ディスクからの OS のブートが可能となる。さらに、CHN 1 1 0 上の CPU 1 1 2 及び I/O プロセッサ 1 1 9 等、管理端末 1 6 0 上の CPU 1 6 1 等、ディスク制御部 1 4 0 上の CPU 1 4 2 等、及び記憶デバイス 3 0 0 とが協働することにより、短時間で効率的な記憶装置システム 6 0 0 の起動、及びソフトウェアのインストールを達成することができる。

【 0 0 7 6 】

本実施の形態は、予め記憶装置システム 6 0 0 に対して CHN 1 1 0 を挿入した後、記憶装置システム 6 0 0 自体の電源を投入する方法について説明したが、本実施の形態はこの場合に限られるものでなく、既に電源の投入されている記憶装置システム 6 0 0 に対してあらたに CHN 1 1 0 を挿入することにより CHN 1 1 0 に電源を投入する方法を取ることも好ましい。

【 0 0 7 7 】

以上本実施の形態について説明したが、上記実施の形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得るとともに、本発明にはその等価物等も含まれる。

【 0 0 7 8 】

【発明の効果】

本発明によれば、複数の異種ネットワークに接続可能なように全く新しく発明された記憶装置システムを提供することができ、さらに、かかる記憶装置システムを発明するにあたり必要とされる記憶デバイス制御装置、及びデバイス制御装置の起動を制御する方法をも提供することができる。

【図面の簡単な説明】

【図 1】 本実施の形態に係る記憶装置システムの全体構成を示すブロック図である。

【図 2】 本実施の形態に係る管理端末の構成を示すブロック図である。

【図 3】 本実施の形態に係る物理ディスク管理テーブルを示す図である。

【図 4】 本実施の形態に係る LU 管理テーブルを示す図である。

【図 5】 本実施の形態に係る記憶装置システムの外観構成を示す図である。

【図 6】 本実施の形態に係る記憶デバイス制御装置の外観構成を示す図である。

【図 7】 本実施の形態に係る CHN のハードウェア構成を示す図である。

【図 8】 本実施の形態に係るメモリに記憶されるデータの内容を説明するための図である。

【図 9】 本実施の形態に係るメタデータを示す図である。

【図 10】 本実施の形態に係るロックデータを示す図である。

【図 11】 本実施の形態に係る CHN 上の CPU と I/O プロセッサとの通信経路を示す図である。

【図 12】 本実施の形態に係る I/O プロセッサから CPU に対してやり取りされるデータを示す図である。

【図 13】 本実施の形態に係る CPU から I/P プロセッサに対してやり取りされるデータを示す図である。

【図 14】 本実施の形態に係る CHN 上の内部 LAN を介したハードウェア構成を示す図である。

【図 15】 本実施の形態に係るディスク制御部を示す図である。

【図 16】 本実施の形態に係る記憶装置システムのソフトウェア構成図である。

【図 17】 本実施の形態に係る記憶装置システムの起動処理及びインストール処理のフローチャートである。

【符号の説明】

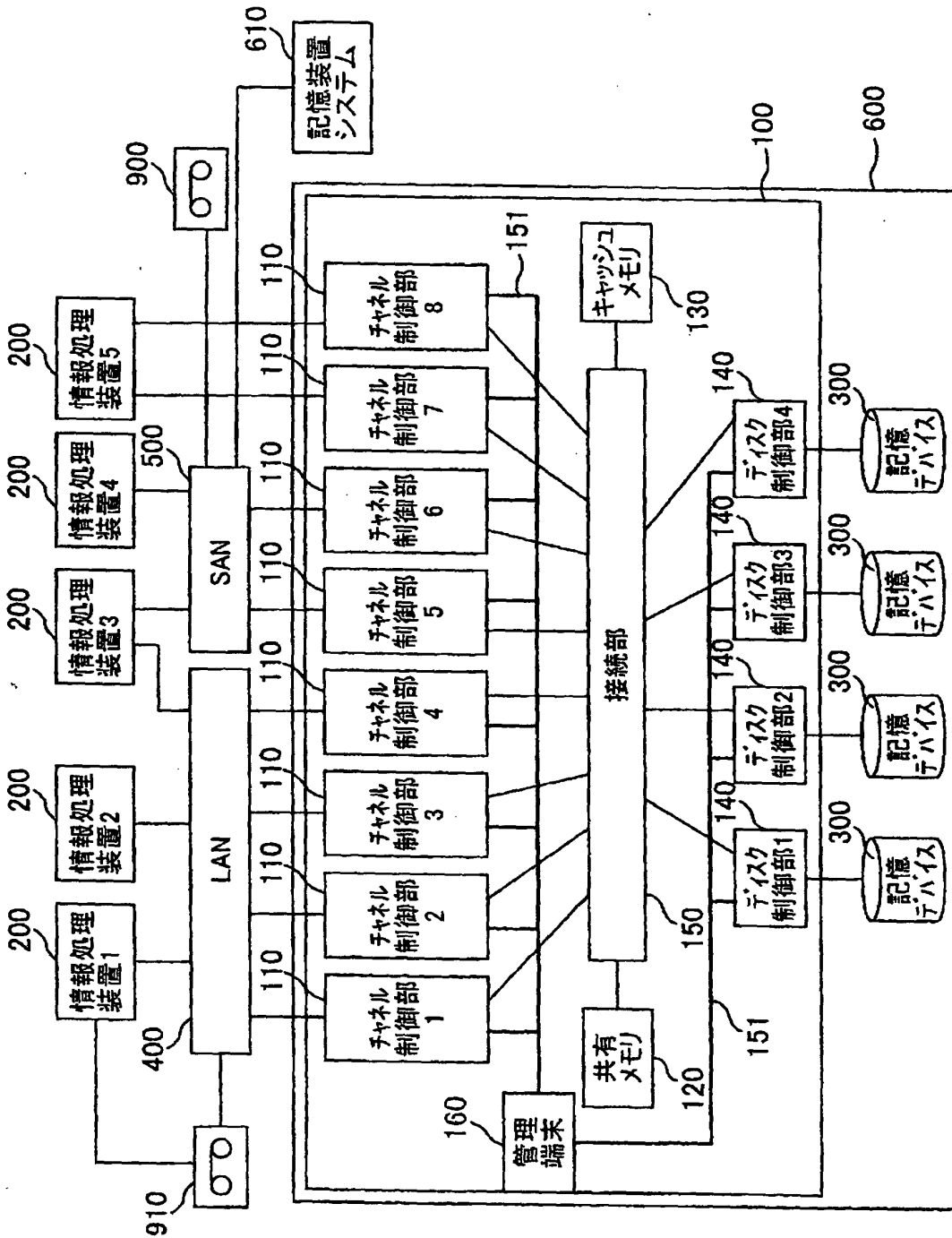
- 1 0 0 記憶デバイス制御装置
- 1 1 0 チャネル制御部
- 1 1 1 ネットワークインタフェース部
- 1 1 2 CPU
- 1 1 3 メモリ
- 1 1 4 入出力制御部
- 1 1 5 NVRAM
- 1 1 6 ボード接続用コネクタ
- 1 1 7 通信コネクタ

- 1 1 8 回路基板
- 1 1 9 I / O プロセッサ
- 1 2 0 共有メモリ
- 1 3 0 キャッシュメモリ
- 1 4 0 ディスク制御部
- 1 5 0 接続部
- 1 5 1 内部 L A N
- 1 6 0 管理端末
- 6 0 0 記憶装置システム
- 8 0 1 B I O S
- 8 0 2 通信メモリ
- 8 0 3 ハードウェアレジスタ群
- 8 0 4 N V R A M

【書類名】 図面

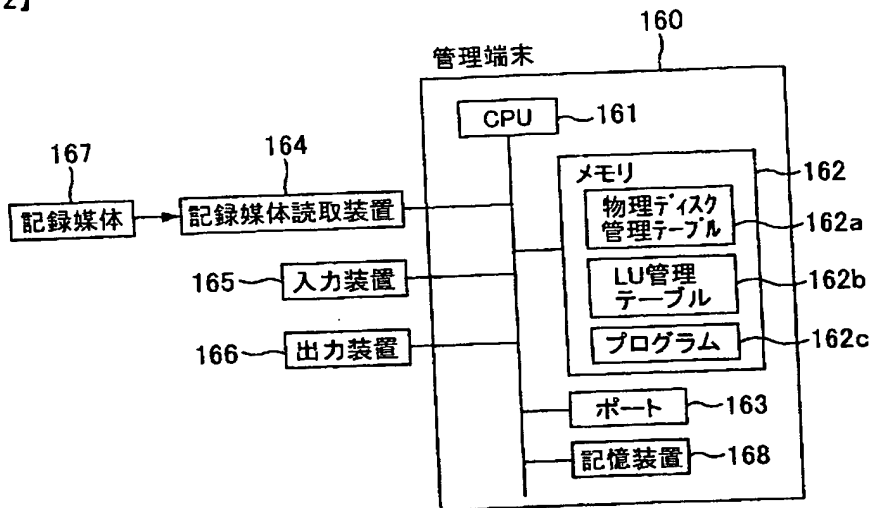
【図1】

【図1】



【図 2】

【図2】



【図 3】

【図3】

162a

物理ディスク管理テーブル

ディスク番号	容量	RAID	使用状況
#001	100GB	5	使用中
#002	100GB	5	使用中
#003	100GB	5	使用中
#004	100GB	5	使用中
#005	100GB	5	使用中
#006	50GB	—	未使用
⋮	⋮	⋮	⋮

【図 4】

【図4】

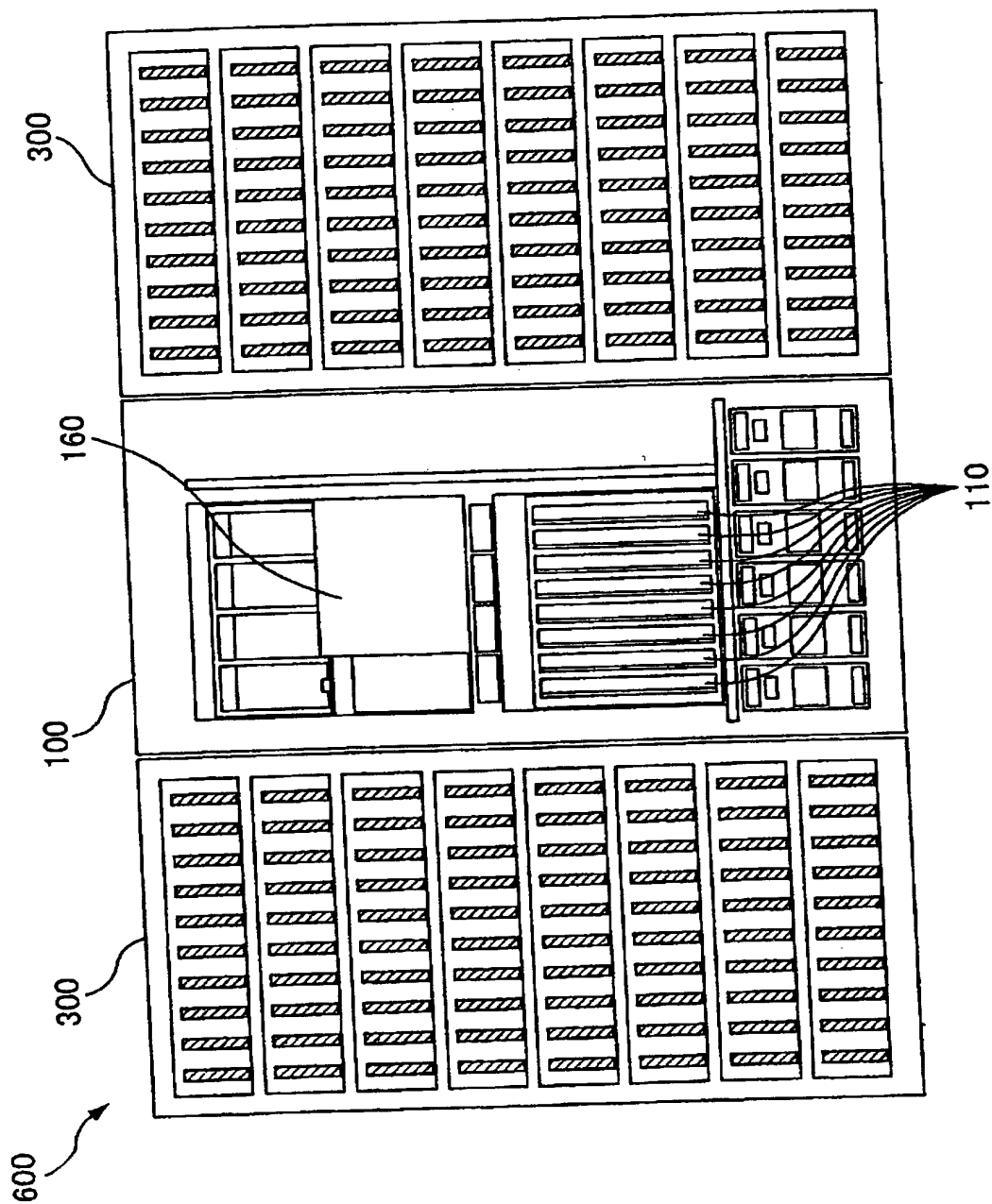
162b

LU管理テーブル

LU番号	物理ディスク	容量	RAID
#1	#001,#002,#003,#004,#005	100GB	5
#2	#001,#002,#003,#004,#005	300GB	5
#3	#006,#007,	200GB	1
⋮	⋮	⋮	⋮

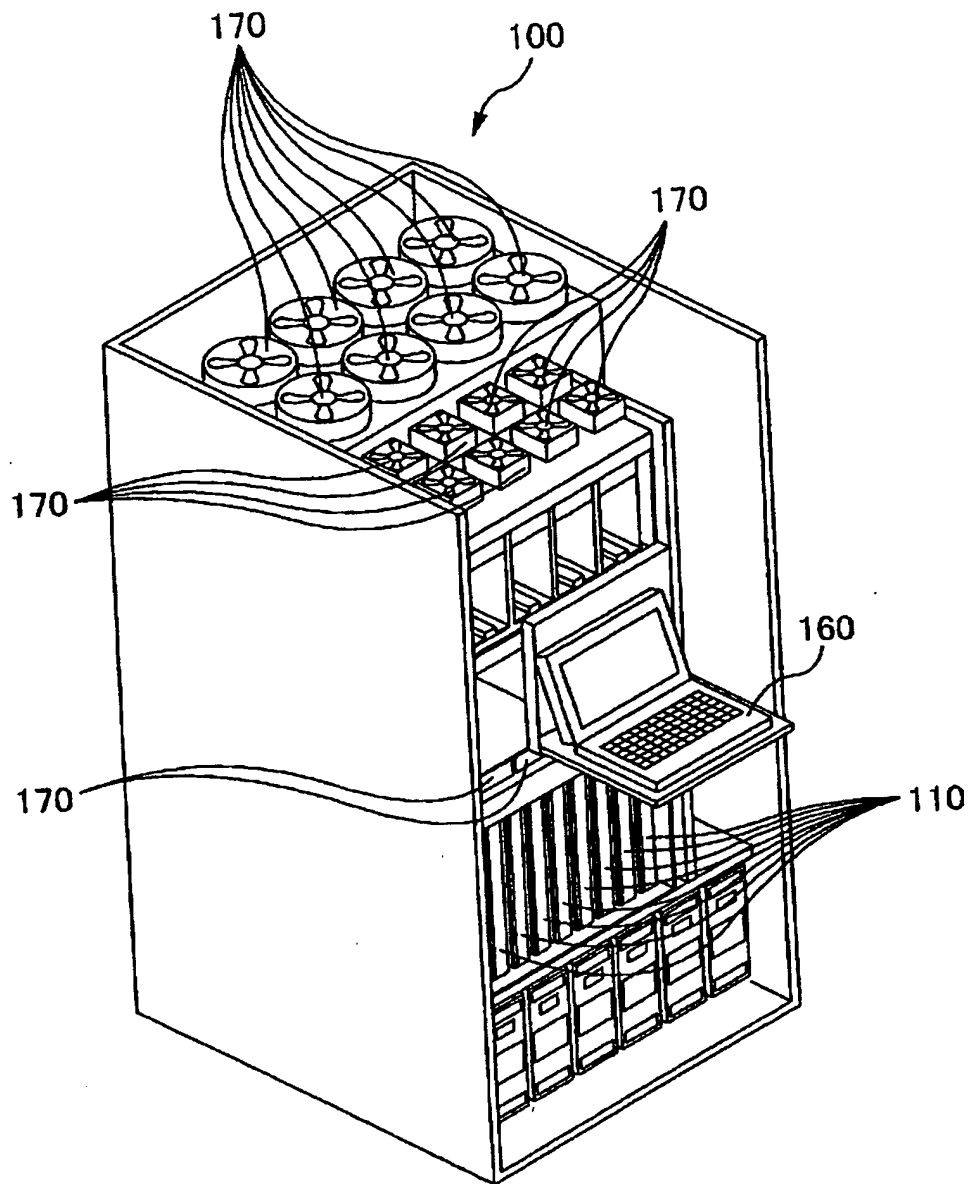
【図 5】

【図 5】



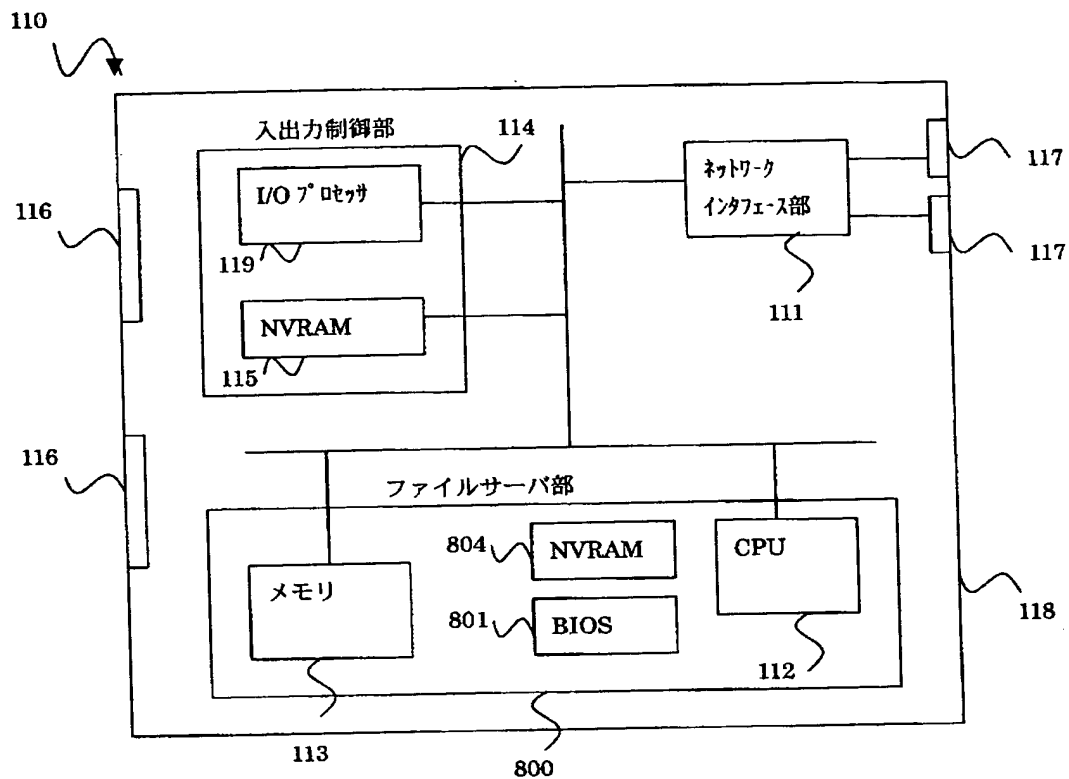
【図6】

【図6】



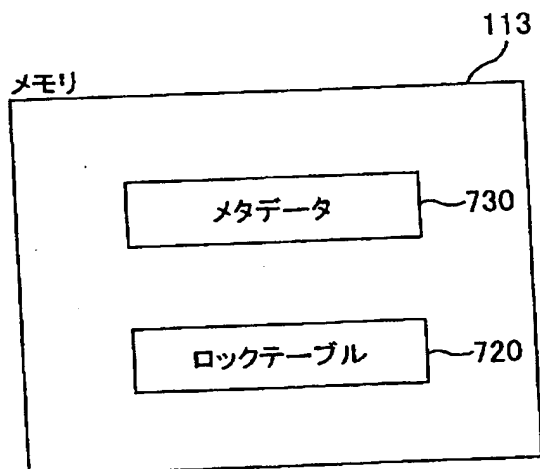
【図7】

【図7】



【図8】

【図8】



【図 9】

【図 9】

730

メタデータ

ファイル名	先頭アドレス	容量	所有者	更新時刻
A	7BSA	200MB	X	0:00
B	05BF	50MB	X	7:57
C	1F30	100MB	Y	9:15
D	470B	100MB	Z	15:20
⋮	⋮	⋮	⋮	⋮

【図 10】

【図 10】

721

ファイルロックテーブル

ファイル名	ロック状態
A	ロック中
B	—
C	—
D	ロック中
⋮	⋮

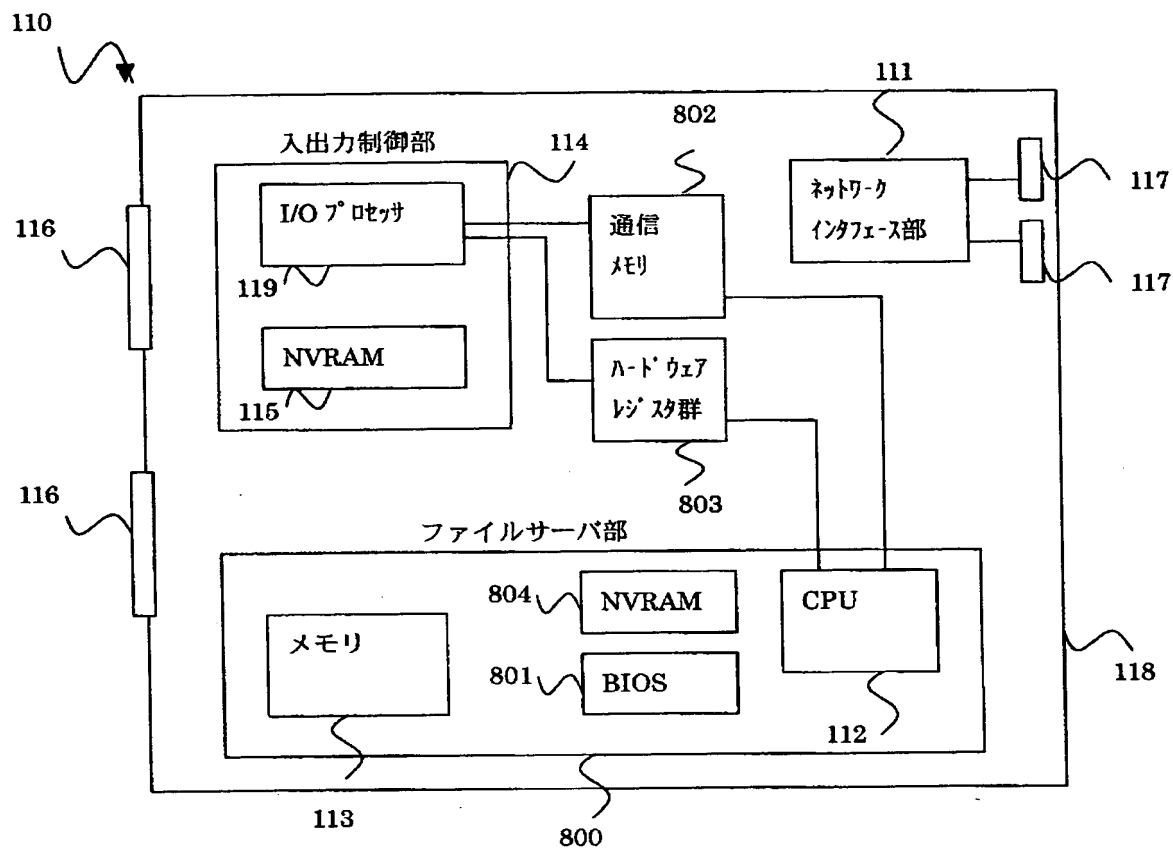
722

LUロックテーブル

LU	ロック状態
共有	—
1	ロック中
2	—
⋮	⋮

【図 1 1】

【図 1 1】



【図 1 2】

【図 1 2】

7-ド	31	24	23	17	16	8	7	0
0	診断実行フラグ		(Reserved)		起動デバイス種別			
1	ドライブ番号 1				ドライブ番号 0			
2	時刻情報							
3	(Reserved)				コマンドリトライ回数		コマンドタイムアウト値	
4	温度情報#3		温度情報#2		温度情報#1		温度情報#0	

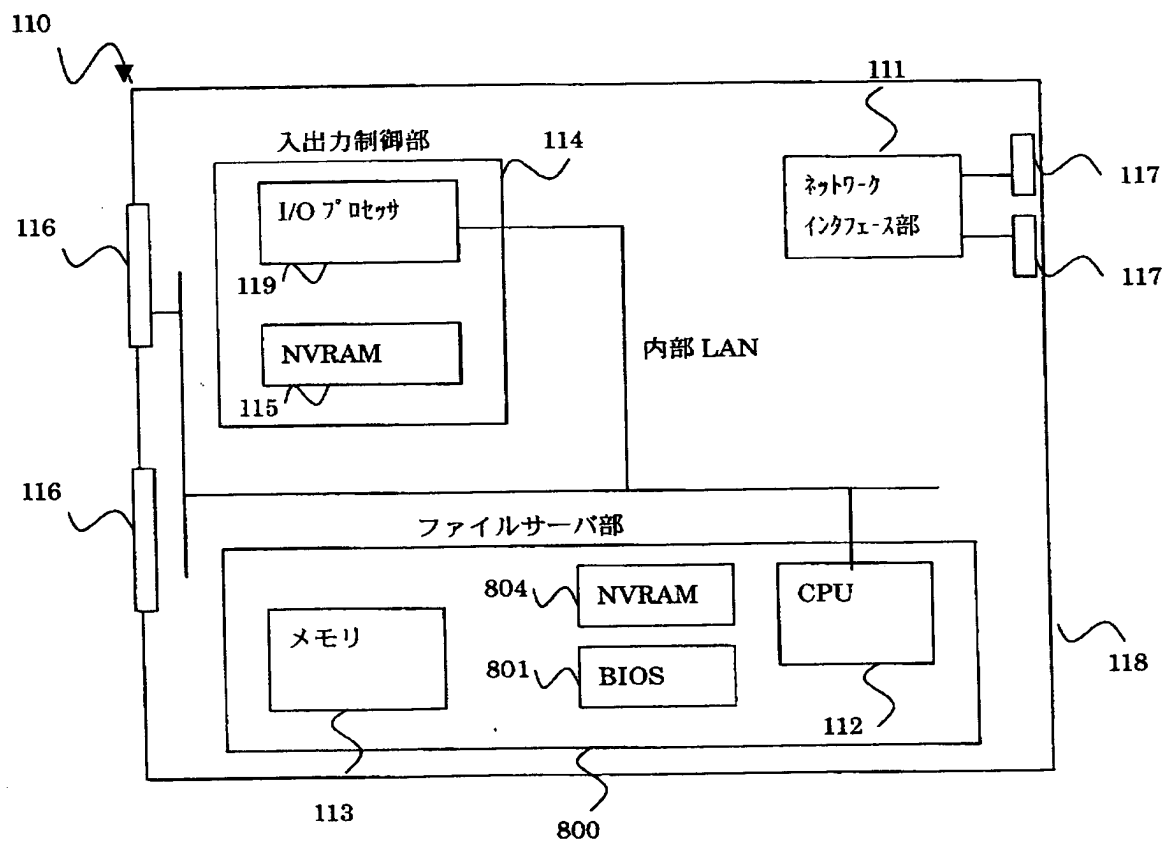
【図13】

【図13】

ワード	31	24	23	17	16	8	7	0
0	MAC アドレス							
1	オペディング 領域							
2	BIOS バージョン							

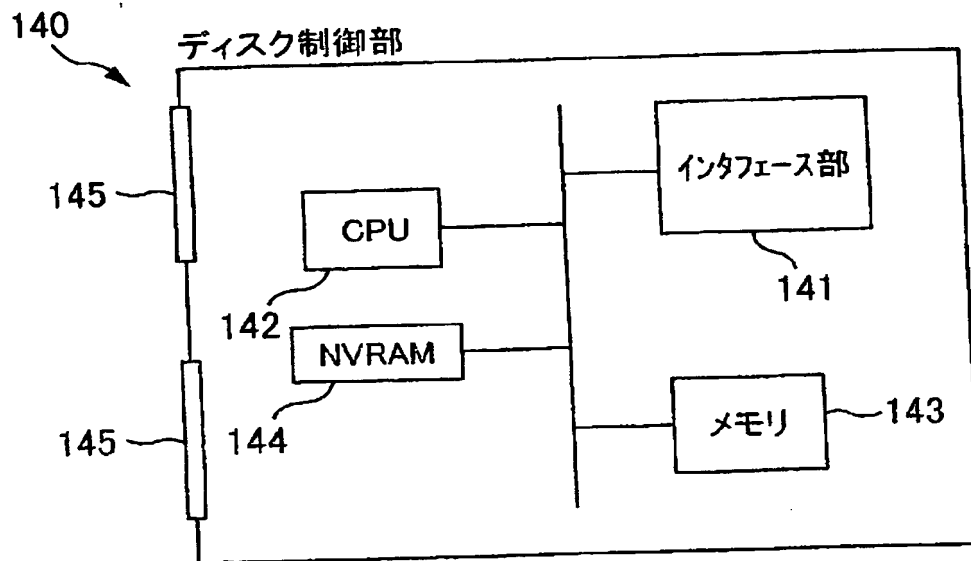
【図14】

【図14】



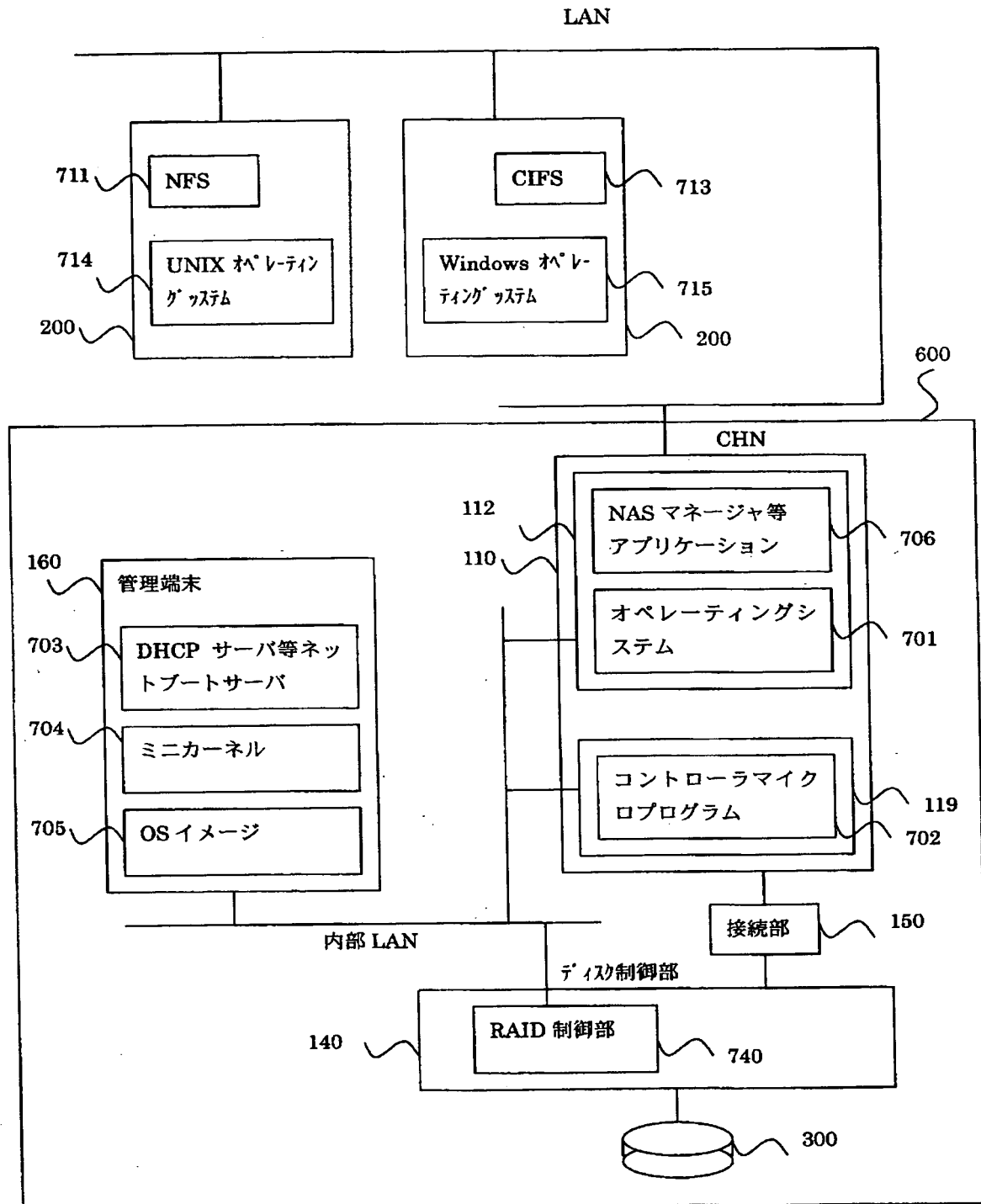
【図15】

【図15】



【図 16】

【図 16】



【図 17】

【図 17】

順番	CPU	I/O プロセッサ	管理端末	記憶デバイス
1	電源投入	電源投入	電源投入	電源投入
2	BIOS 起動	ハードウェア初期化	端末起動	ドライブスビンアップ開始
3	I/O プロセッサ指示待ち			
4	ハードウェア診断開始	← 診断開始要求		
5	ハードウェア診断終了			
6	MAC アドレス格納	→		
7	電源再投入	← CPU リセット		
8	BIOS 起動			ドライブスビンアップ完了
9	I/O プロセッサ指示待ち			
10			MAC アドレス取得	
11			リトライ回数・リトライタイムアウト値・起動デバイス種別(ネットアダプタ)設定	
12		ドライブスビンアップ判定		→
13		ドライブスビンアップ完了検出	←	
14		温度情報・リトライ回数・リトライタイムアウト値・診断実行フラグ(診断ステップ)・起動デバイス種別(ネットアダプタ)格納		
15	CPU 処理続行指示検出	← CPU 処理続行指示		
16	温度情報・リトライ回数・リトライタイムアウト値・診断実行フラグ・起動デバイス種別取得			
17	ネットアダプタ開始			
18	ネットアダプタ要求発行		→ ネットアダプタ要求受領	
19	OS インストール開始(ネットクライアント)		OS 転送(ネットサーバ)	
20	OS イメージ転送			→
21	インストール完了通知	→ インストール完了通知検知		
22	電源再投入	← CPU リセット		
23	BIOS 起動			
24	I/O プロセッサ指示待ち			
25		時刻情報・起動デバイス種別(ディスク)・ドライブ番号格納		
26	CPU 処理続行指示検出	← CPU 処理続行指示		
27	時刻情報取得			
28	ディスクアダプタ開始			
29	OS ロード	←		
30	OS 起動			
31	通常動作			

【書類名】 要約書

【要約】

【課題】

複数の異種ネットワークに接続可能な記憶装置システムを提供するとともに、かかる記憶装置システムを発明するにあたり必要とされる記憶デバイス制御装置、及びデバイス制御装置の起動を制御する方法を提供する。

【解決手段】

本発明の記憶装置システムは、情報を格納する複数の記憶装置と、前記複数の記憶装置に対する情報の格納を制御する記憶装置制御部と、前記記憶装置制御部に接続される接続部とを有し、さらに、前記接続部を介して前記記憶装置制御部に接続されるとともに、自記憶装置システムの外部の第一のネットワークに接続され、前記外部の第一のネットワークを介して受けた第一の形式の情報を第二の形式の情報に変換して前記複数の記憶装置へのアクセスを要求される第一のプロセッサと、前記第一のプロセッサからのアクセス要求に応じて前記接続部及び前記記憶装置制御部を介して前記複数の記憶装置へアクセスするとともに、前記第一のプロセッサの起動を制御する第二のプロセッサとを有する第一の通信制御部とを有する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2003-015525
受付番号	50300108581
書類名	特許願
担当官	第七担当上席 0096
作成日	平成15年 1月27日

<認定情報・付加情報>

【提出日】	平成15年 1月24日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台4丁目6番地
氏 名	株式会社日立製作所